Lab 1 Practice: Introduction to RStudio

Background

In this lab, you will explore R.

Practice

1. Open RStudio. In the lower right hand pane, click on the tab that says "Packages". Then click on "Install", and type in `knitr`.
2. Open a blank `R markdown` document using the green circle plus icon in the top left hand corner of RStudio. This is where you should type and save your code and answers. Write "Lab 1 Practice" in the title and include names of all present group members at the top of the document.
3. Execute the following calculation in R (your result should be 1.57). Write your code inside of an r 'chunk' in the markdown document. This is to demonstrate that R can be used as a calculator and illustrate the importance of order of operations. Some arithmetic symbols you can use in R code include: `() + - / * sqrt()`. (This calculation takes the form of a *test statistic*, which we will learn later in the quarter.)

   $$\frac{3.5 - 3.3}{0.9/\sqrt{50}}$$
4. Import the `yrbss2013` data set. Save the code from your `History` tab that you used to import the data set into an R chunk.
5. View a summary of the data set using the command below (make sure to put it in a chunk!). Examine the `sad` variable, and report how many individuals responded "yes". Interpret what this means using the codebook on page 2.

```
summary(yrbss2013)
```

6. Create a scatterplot with `height_m` on the x-axis and `weight_kg` on the y-axis. Modify the plot to use a point type of filled in cirlces with `pch=19` and select the color of your choice instead of black (*http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf*). Include the modified scatterplot in your lab report. How would you generally describe the relationship between height and weight?

```
plot(yrbss2013$height_m,yrbss2013$weight_kg,pch=1,col="black")
```

7. Using a single variable of your choice from the `yrbss2013` data set, try a new R command of your choice from the "Quick R Commands" pdf on PolyLearn. Execute the command and interpret your result.
8. Submit this assignment compiled as an html file from R markdown. Make sure each response is labeled with the corresponding question numbers.

The Data: Youth Risk Behavior Surveillance System

The Youth Risk Behavior Surveillance System (YRBSS) has been conducted every two years since 1991 by the Centers for Disease Control and Prevention (CDC) in order to obtain information from adolescents regarding trends in risky behavior, such as smoking, drinking, drug use, diet, and physical activity. In 2013, 47 states participated in this school-based survey, yielding 13,583 respondents and 213 variables. Full survey and data documentation can be accessed on the CDC website. A subset of this data set which has no missing data for 17 selected variables is provided in the file `yrbss2013.csv`[1].

`age`	*Q1: How old are you?*
`gender`	*Q2: What is your sex?*
`height_m`	calculated variable: height in meters
`weight_kg`	calculated variable: weight in kilograms
`bmi`	calculated variable: body mass index=`weight_kg`/`height_m`^{2}
`BMIPCT`	calculated variable: BMI percentile for age and sex
`seatbelt`	*Q9: How often do you wear a seat belt when riding in a car driven by someone else?*
`seatbelt2`	calculated variable: `seatbelt` never vs otherwise
`ride_drunkdriver`	*Q10: During the past 30 days, have you ridden in a car or other vehicle driven by someone who had been drinking alcohol?*
`drive_drunk`	*Q11: During the past 30 days, how many times did you drive a car or other vehicle when you had been drinking alcohol?*
`drive_text`	*Q12: During the past 30 days, on how many days did you text or e-mail while driving a car or other vehicle?*
`carried_weapon`	*Q13: During the past 30 days, did you carry a weapon such as a gun, knife, or club?*
`unsafe_school`	*Q16: During the past 30 days, did you not go to school because you felt you would be unsafe at school or on your way to or from school?*
`bullied`	*Q24: During the past 12 months, have you ever been bullied on school property?*
`sad`	*Q26: During the past 12 months, did you ever feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing some usual activities?*
`days_smoke`	*Q33: During the past 30 days, on how many days did you smoke cigarettes?*
`days_drink`	*Q43: During the past 30 days, on how many days did you have at least one drink of alcohol?*

[1]The variables `days_smoke` and `days_drink` were originally coded in categories of '0 days', '1 or 2 days', '3 to 5 days', '6 to 9 days', '10 to 19 days', '20 to 29 days', and 'All 30 days'. The values provided in this data set were randomly generated according to the category specified.